结合学习率调整的自适应特征融合相关滤波跟踪算法 *

成　悦，李建增

(陆军工程大学石家庄校区 无人机工程系，石家庄 050003)

摘　要：针对单一特征存在的缺陷和目标快速变化时易跟丢的问题，提出了一种结合学习率调整的自适应特征融合相关滤波跟踪算法。算法采用互补的梯度特征和颜色特征进行特征融合，通过计算滤波响应的大小来决定下一帧在融合特征中各自所占的权重，凸显优势特征，使目标与背景更具区分度。提取目标后需要更新滤波器，为了避免滤波器跟不上目标变化的情况发生，引入学习率调整机制，使滤波器更新速度能够随目标外观变化进行在线调整。因此，相较同类特征融合算法，本算法准确高效，且对于快速形变目标的鲁棒性更强。实验证明，本算法在精度和成功率上都比现有相关滤波算法更优，具有一定的应用价值。

关键词：目标跟踪；相关滤波；特征融合；自适应加权；学习率

中图分类号：TP391.4　　doi: 10.3969/j.issn.1001-3695.2018.01.0121

Adaptive feature fusion correlation filter tracking algorithm combined with learning rate adjustment

Cheng Yue, Li Jianzeng

(*Dept. of UAV Engineering*, *Army Engineering University*, *Shijiazhuang* 050003, *China*)

Abstract: In view of the defects of single feature and the problem of easy to miss when the target is fast changing, this paper proposes an adaptive feature fusion correlation filter tracking algorithm based on learning rate adjustment. The algorithm used complementary gradient features and color features to do the feature fusion, and decides the weights of the features in the fusion feature in next frame by calculating the size of the filter response of each feature, so as to highlight the dominant features and make the target and background more discriminative. After extracting the target, we need to update the filter. In order to avoid the situation that filter can not keep up with the change of the target, the learning rate adjustment mechanism is introduced, so that the update speed of the filter can be adjusted online with the appearance of the target. Therefore, compared with the similar feature fusion algorithm, our algorithm is more accurate and efficient, and the robustness of the fast deformation target is stronger. The experiment shows that our algorithm is better than the existing correlation filter algorithms in accuracy and success rate, and has a certain application price.

Key words: target tracking; correlation filtering; feature fusion; adaptive weighting; learning rate

0　引言

视觉目标跟踪技术是计算机视觉中最活跃的研究课题之一，它在各种运动分析、行为识别、监控及人机交互等方面都发挥着重要作用。目标跟踪技术的主要任务就是在图像序列中估计目标轨迹，并能够准确地在各种外界或内部的干扰中稳定地锁定目标的位置。近几年，目标跟踪技术涌现出了很多性能和速度都很优秀的算法，其中相关滤波类算法是目前较先进的一类算法，受到广泛关注和研究。

相关滤波器最早是在 Bolme 等人[1]提出的 MOSSE 算法中进行应用的，此后 Henriques 等人[2]发掘出相关滤波器的优势，在 2012 年的 ECCV 会议上发表了一篇非常高效的运动跟踪文章，也就是后来的 exploiting the circulant structure of tracking-by-detection with kernels（CSK）算法，获得了较好的效果。随后，他们又提出了经典的 kernel correlation filter (KCF)[3]算法和 discriminative correlation filter(DCF)算法，成为了后续众多算法的基础。CSK 算法采用的是灰度特征，KCF 算法采用的是 histogram of oriented gradient(HOG)特征，而此后 Danelljan 等人[4]提出的基于颜色名称的 color name(CN)算法，把颜色信息细化为了 11 个维度并作为特征进行相关滤波，达到了较好的精度。文献[5]采用了颜色特征与纹理特征相融合的方法，文献[6,7]也是采用了多种特征融合的方式对目标进行描述，取得了较好

收稿日期：2018-01-31；**修回日期：**2018-03-20　**基金项目：**国家自然科学基金资助项目（51307183）

作者简介：成悦（1993-），女，河北省石家庄人，硕士研究生，主要研究方向为视频目标跟踪及图像处理（18633063191@163.com）；李建增（1966-），男，副教授，硕士，主要研究方向为图像处理、情报传递.

的效果。除了特征方面的改进外，以下文献从各个角度对相关滤波进行了改进。SRDCF[8]和 DeepSRDCF[9]是基于 DCF 的改进算法，二者采用的特征不同，后者检测区域更大，在效果上也有一定程度上的提升。文献[10]采用了基于相位特征的高斯核相关算子，改善了跟踪器对光照的敏感度。文献[11]考虑到了时空相关性，以此来构建目标空域滤波模型。近几年，神经网络方法以其精度好、智能程度高的优势逐渐受到人们的关注。文献[12~14]将神经网络与相关滤波相结合，同样达到较好的效果。

在近几年的比较先进的相关滤波算法中， discriminative scale space tracking（DSST）[15]算法虽然加入了尺度变换，但采用的是单一特征，引入较多噪声，导致跟踪不稳。SRDCF 算法鲁棒性很好，但速度很慢，无法满足实时性。Sum of template and pixel-wise learners(Staple)[16]算法虽然采用了 HOG 特征结合颜色特征的方法，但采用的是经验值权重，无法针对目标及环境的变化自主调节，适应性较差。C-COT[17]算法采用了神经网络来提取特征，极大地增加了计算复杂度，速度较慢。针对以上算法存在的单一特征具有局限性和无法满足实时性的问题，本文对特征的合理提取与融合策略进行了改进，提取优缺点互补的 HOG 特征和 CN 特征，并通过二者的滤波响应图来自适应调整融合特征中二者所占比例，使特征能够更稳定地用于跟踪。

在提取目标后，滤波器的更新同样影响着跟踪效果，文献[18]采用了平均差来计算图像的变化率，准确度较低，本文通过引入欧式距离来解决这一问题，将更新系数由一常量变为参数，使滤波器模型能够根据目标变化快慢而自行调整。最后进行了多组实验，验证了本算法的有效性。

1　相关滤波算法基本原理

相关滤波算法主要通过训练相关滤波器得到响应图，利用循环矩阵对角化的特征和快速傅立叶变换，使滤波过程的计算变得快速高效。相关滤波类算法主要包括滤波器更新和目标检测两个部分。

1.1　相关滤波器的训练和更新

相关滤波器采用的是密集采样，即得到上一帧的目标后，对其进行行方向和列方向上的循环移位，以达到密集采样的效果，如图 1 所示。

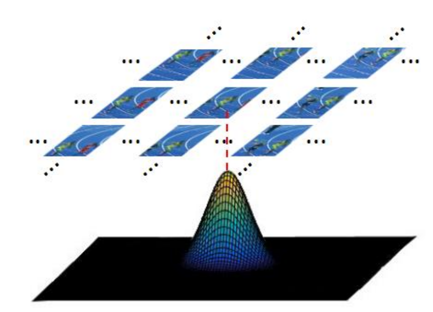


图 1　循环采样对应高斯函数示意图

得到样本后就需要对滤波器进行训练，文献[1]中采用的是岭回归方法，目标函数为正则化项和损失函数相加，如式(1)所示。

$$w=\min_{w}\sum_{i}\left(f(x_i)-y_i\right)^2+\lambda\|w\|^2 \tag{1}$$

在非线性情况下引入核函数，将相关滤波的主要变化参数变为滤波器系数 α 和目标模型 x，如式(2)~(4)所示。

$$w=\sum_{i}\alpha_i\Phi(x_i) \tag{2}$$

$$\alpha=Fdiag(\hat{K}^{xx}+\lambda)F^{H}y \tag{3}$$

$$\hat{\alpha}=\frac{\hat{y}}{(\hat{K}^{xx}+\lambda)^{*}} \tag{4}$$

其中：$\hat{\alpha}$ 为傅里叶域中的 α。相关滤波器的更新方式主要是线性插值方法，设置插值系数为 η，如式(5)所示。

$$\begin{cases}\alpha_t=(1-\eta)\times\alpha_{t-1}+\eta\times\alpha_t{}'\\ x_t=(1-\eta)\times x_{t-1}+\eta\times x_t{}'\end{cases} \tag{5}$$

1.2　目标检测

训练得到当前帧的滤波器模型后，就需要对新一帧的目标位置进行检测。待检测样本集是预测区域和由其移位得到的样本集合，单个样本和一次计算所有样本的响应分别如式(6)(7)所示。

$$f(z_j)=\alpha^{T}\Phi(X)\Phi(z_j) \tag{6}$$

$$\hat{f}(z)=\hat{K}^{xz}\times\hat{\alpha} \tag{7}$$

其中得到的 $f(z)$ 为与 z 相同大小的滤波响应图，图中概率最大的点即视作目标在新一帧中的位置。并获取新一帧的滤波器模型和主要参数进行训练更新，以做好对下一帧目标进行检测的准备。

相关滤波器的主要工作流程如图 1 所示。

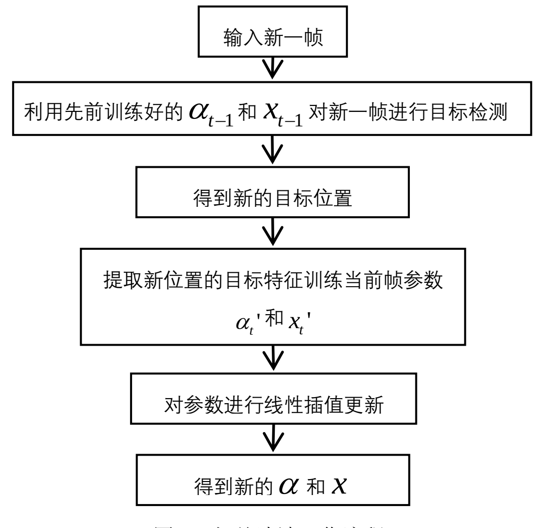


图 1　相关滤波工作流程

2　基于响应图的自适应特征融合

特征的提取是目标跟踪的基础和前提。找到一种好的特征提取方式能够最大程度地区分目标和背景，从而更好地提升算法的跟踪效果。

传统的算法一般采用单一的特征进行跟踪，但是每个特征都会存在缺点，当遇到不易处理的情况时就会导致跟踪效果下降。近年较先进的几种算法所选用的特征一般为颜色特征、HOG 特征和深度特征，因为这三种特征的区分度好，能够更准确地描述目标，但它们都存在各自的优缺点。深度特征一般采用神经网络进行提取，越精准的网络层数越多，计算也就越复杂，基本无法满足实时性。颜色特征具有旋转不变性，且不受目标物大小和形状的变化影响，但是当光照剧烈变化时，单一的颜色特征就不能够很好地分辨出目标的位置，从而导致目标丢失。而 HOG 特征通过计算和统计图像局部区域的梯度方向直方图来构成特征，保持了几何和光学转化的不变性，但遇到目标形状快速变化的情况时 HOG 特征容易跟不上目标的变化从而丢失目标。由此可见，在较为常用的特征中，颜色与梯度这两个特征是互补的，将二者结合起来不仅可以很好地弥补各自的缺点，使跟踪能够适应各种情况，而且仍能保证实时的跟踪速度。

但简单的线性叠加并不能完全达到理想效果，故本文提出了一种根据滤波响应图来对特征进行自适应加权的特征融合方法，当一种特征效果变差时，自适应地降低该特征的权重并提升另一种特征的权重，使优势特征能够占据主导地位，以更好地处理光照、快速形变等问题。

传统的相关滤波类算法是用岭回归方法来对滤波器进行训练的，它的损失函数如式(8)所示。

$$l=\sum_{i,j}(R_{i,j}-\overline{R_{i,j}})^2+A \tag{8}$$

其中：R 为样本 (i,j) 的滤波响应值，$R_{i,j}$ 为当前帧中样本在滤波器中的实际响应值，$\overline{R_{i,j}}$ 为该样本在滤波器中的理想响应值，二者差距越小证明该样本的滤波响应越大，即滤波器越理想。A 为正则项。因此，如何最小化损失差 $\sum_{i,j}(R_{i,j}-\overline{R_{i,j}})^2$ 是提升跟踪效果的重点。

本算法将损失差的大小作为衡量下一帧特征权重的标准，首先分别计算 HOG 特征和 CN 特征在第 t 帧的损失差：

$$D_t^{CN}=sum(R_t^{CN}-\overline{R_t}^{CN})^2 \tag{9}$$

$$D_t^{HOG}=sum(R_t^{HOG}-\overline{R_t}^{HOG})^2 \tag{10}$$

其中：$sum(\cdot)$ 代表矩阵对应项相加，两个特征各自的自适应权重由下式计算：

$$\omega_t^{CN}=\frac{D_t^{HOG}}{D_t^{CN}+D_t^{HOG}},\quad \omega_t^{HOG}=1-\omega_t^{CN} \tag{11}$$

各自权重的更新策略为

$$\tilde{\omega}_t=(1-\lambda)\times\omega_{t-1}+\lambda\times\omega_t \tag{12}$$

其中：ω_{t-1} 为上一帧更新后用于本帧特征融合的权重，$\tilde{\omega}_t$ 为根据当前帧和先前的结果经过更新之后求出的特征权重，λ 为更新系数。初始 $\omega^{CN}=\omega^{HOG}=0.5$。

分别得到 CN 特征和 HOG 特征的融合权重之后，在滤波响应图层面对两特征进行加权融合，得到最终的融合特征：

$$R_t=\omega_t^{CN}\times R_t^{CN}+\omega_t^{HOG}\times R_t^{HOG} \tag{13}$$

将此融合特征作为最终特征输入滤波器，求得的响应最大处即为目标当前所在位置。

图 2 为本文特征融合方法在 OTB-50[19]标准数据集中的 Bolt 视频序列上运行的两特征权重变化图。图中可以看出，刚开始起跑的时候，由于背景中黄色标志物对目标的颜色干扰，使得 CN 特征权重下降，随着跑离标志物，CN 特征有所回升，但草坪和广告牌的颜色干扰使得 CN 的权重再次下降。此后运动员开始进入加速、匀速和冲刺的阶段，随着运动员速度的增加，边界轮廓变得越来越模糊，因此 HOG 特征权重响应下降，启用 CN 特征进行跟踪，直至最后冲过终点一段距离后两种特征的权重才趋于相等。

图 3 分别为颜色和 HOG 特征在第 60 帧和第 270 帧时的滤波响应图，左侧为 HOG 特征，右侧为颜色特征，因为 HOG 特征提取时以 cell 为单位，而颜色特征是以像素为单位，所以图中 HOG 特征的分辨率低于颜色特征。图 3 中第 60 帧，HOG 特征滤波响应较大，而颜色特征响应较小，因此在图 2 的权重图中，HOG 特征权重较大，颜色特征权重较少。第 270 帧则相反。

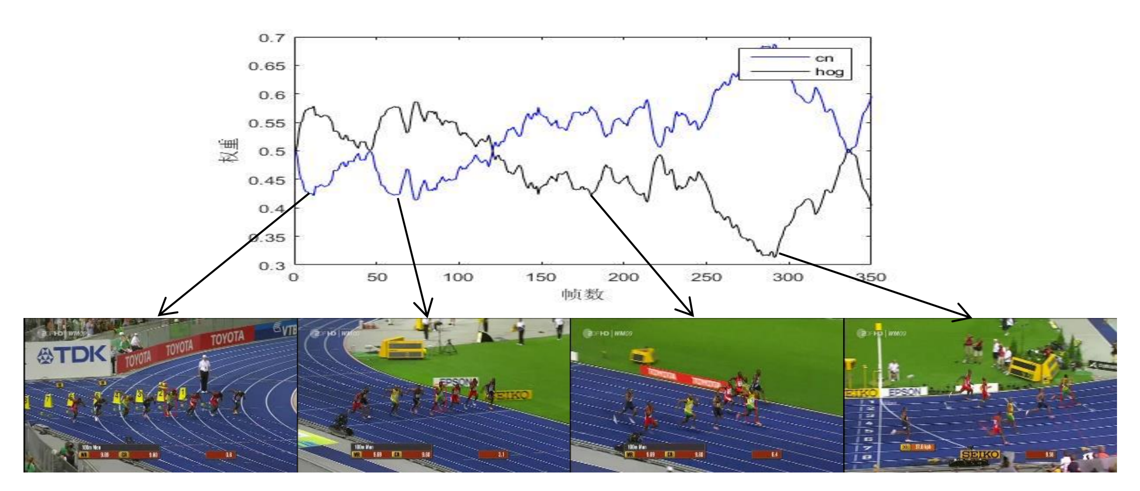
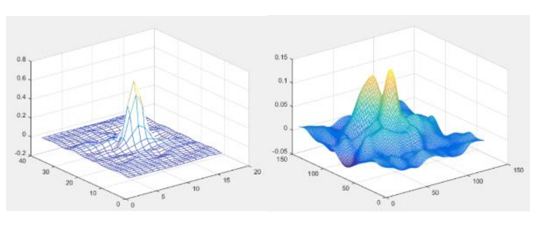


图 2 特征权重变化图



(a) 第 60 帧　　(b) 第 270 帧

图 3 两种特征滤波响应图对比

3　学习速率在线调整

上一章提到，在相关滤波算法中，滤波器参数的更新采用到了更新系数 λ，将以往的滤波器模型与新一帧的模型进行加权融合，得到最新的滤波器模型用于下一帧的目标检测，即更新系数 λ 越大，学习速率越快。一般情况下，更新系数 λ 是一个经验值常量，由实验者在跟踪开始之前进行设定，跟踪过程中不再进行更改。这就导致模型更新的速率可能会跟不上或者超出目标外观更新的速率，在目标变化大时无法准确检测出目标，而目标基本不变时丢失较多以往滤波器模型造成模型单一或过拟合。因此本文对更新系数 λ 进行了改进，使之不再是一个单一的经验值常量，而能够随着目标外观变化的快慢而在跟踪过程中进行调整，以适应不同的目标跟踪情况。

欧氏距离也称欧几里得度量，是一个通常采用的距离定义，它代表在 m 维空间中两个点之间的真实距离。本文中采用欧式距离来计算两帧图片之间的变化量，来判定目标外观是否出现较大变化，并据此对更新系数 λ 的大小进行调整。

对于宽高分别为 M 和 N 的图像块，欧氏距离 d 的计算如式(14)所示。

$$d=\sqrt{\sum_{i,j}^{M,N}(J_{i,j}^{n}-J_{i,j}^{n-1})} \qquad (14)$$

其中：$J_{i,j}^{n}$ 表示第 n 帧的图像块中像素点 (i,j) 的三个颜色通道分别计算距离后取平均值。对其进行归一化：

$$g=\frac{d}{M\times N} \qquad (15)$$

g 越大，表示前后两帧的目标区域差异越大，需要同时加快滤波器的更新速度，即增大 λ；反之则表示目标差异小，需要适当降低滤波器的更新速度，即减小 λ。由此可见，λ 与 g 为正相关关系：

$$\lambda=\eta\times g+\varepsilon \qquad (16)$$

其中：η 为一经验值常量，当目标完全不变时，取 $\lambda=\varepsilon$=0.001。

图 4 为对更新系数 λ 进行改进前后在目标快速变化的 Ironman 视频序列上进行测试时的对比图，可见在线对 λ 调整能够更准确地识别快速变化的目标。

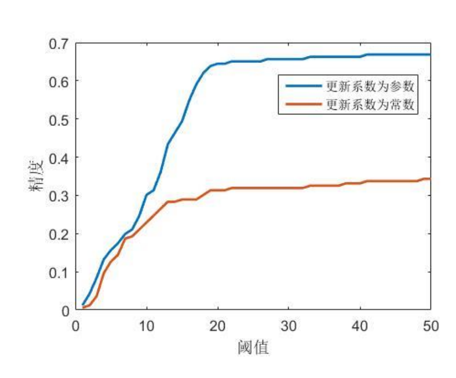
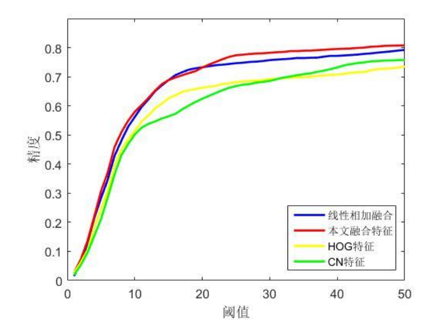


图 4 更新系数 λ 采用本文改进方法前后精度对比图

4　实验验证

本文所进行的实验采用的平台为 MATLAB R2015b。实验视频选用目前国际会议通用的标准数据集 OTB-50 中的视频序列，包括 Bolt、Soccer、Football、Shaking、Skating1、Deer、Ironman、Diving、Singer2、Tiger2、Panda 等，涵盖了光照变化、颜色变化、快速移动、遮挡、尺度变化等一系列视频目标跟踪的过程中常见的问题，具有较高的代表性。本文首先对于采用 HOG 特征、CN 特征、两特征直接线性相加融合和本文特征融合方式进行对比，然后将本文算法与几种现有的常见算法进行比较。测评准则采用了常见的精度和正确率曲线来对跟踪进行评价，精度即实际跟踪框中心坐标与标注的目标框中心坐标之间的距离，正确率主要看实际跟踪框与标注目标框之间面积的重叠率，即重叠面积与总面积之比。精度曲线横轴为距离阈值，纵轴是距离超过当前阈值的帧数与总帧数的比值。正确率曲线的横轴是重叠率阈值，纵轴是重叠率大于当前阈值的帧数与总帧数的比值。

图 5 是采用不同特征进行跟踪的效果对比图。可以看出，融合特征在整体上是优于单独使用某一特征的，但融合方式的不同也导致了最后效果的不同。本文算法寻找优势特征，加大优势特征权重，使跟踪效果明显优于两特征直接线性叠加的融合方式。



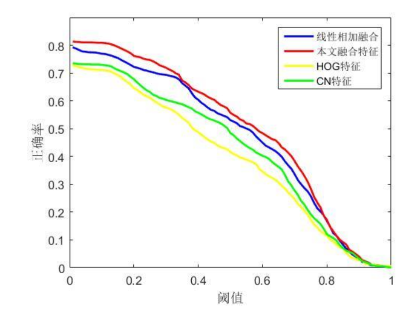
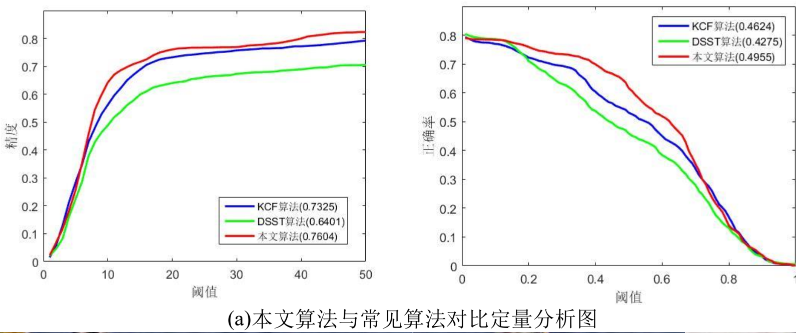
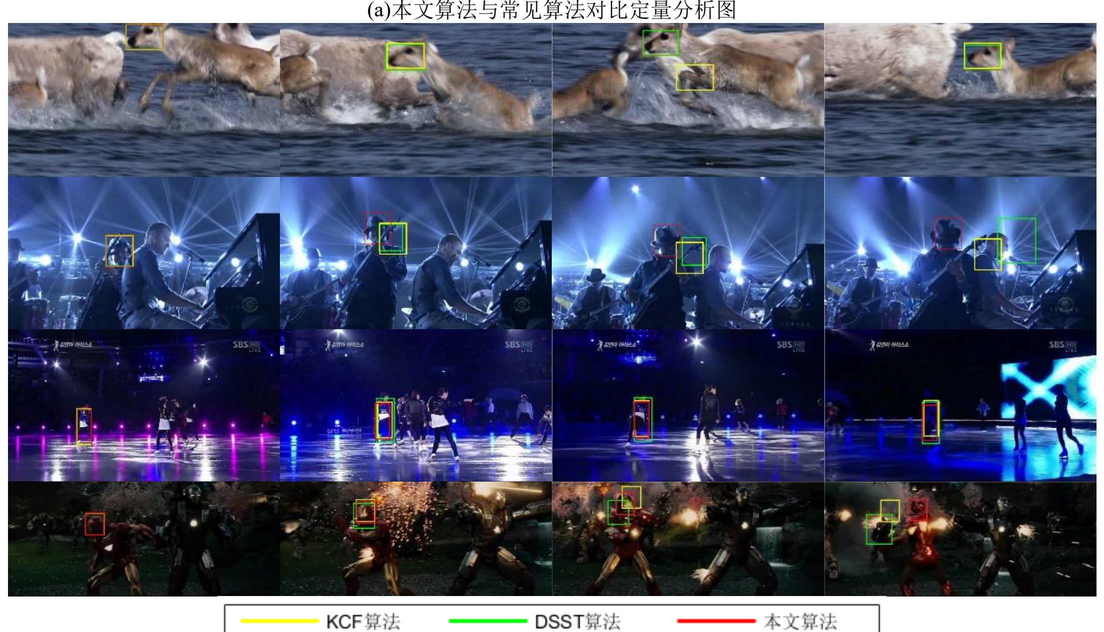


图 5　不同特征跟踪效果对比图

图 6 为本文算法与现有常见算法进行的对比实验结果图。由图可见，本文算法相比其他两种算法具有更高的精度和正确率，且能够满足实时性。



(a)本文算法与常见算法对比定量分析图



(b)本文算法与常见算法对比定性分析图

图 6　本文算法与常见算法对比图

通过多次实验取均值的方法进行统计得到最后的结果，证明了本文算法针对视频中出现的快速移动、光照变化和快速形变等难点具有较好的处理能力，在对比实验中取得了较好的效果。在速度方面，由于本文对特征自适应地分配权重，需要分别计算两种特征的滤波响应并进行加工处理，同时在每次迭代中都加入了图片变化量的计算，增加了计算复杂度，但相关滤波器本身采用了快速傅立叶变化进行提速，因此本文算法速度相较 KCF 算法虽有所下降，但仍能满足对实时性的要求。工程上认为帧率大于 25 fps 就可用于实时跟踪，本文算法达到 35 fps 左右，可以用于视频监控、航拍等需要实时跟踪的多个领域。

5　结束语

本文将自适应调整权重的特征融合方法与更新系数的改进方法结合起来，应用在相关滤波的跟踪框架下，结合快速傅里叶变化的速度优势和互补特征融合的高区分度，达到了一个速度与精度兼优的跟踪效果。相比其他自适应特征融合算法，本文算法合理选择互补特征进行融合，在提升精度的同时减少时间损耗，采用损失差作为分配权重的参数，简洁准确，且加入了更新系数的调整，提升了目标快速变化情况下的鲁棒性，使算法具备实际应用于工程的条件。最后进行了对比实验，结果证明本文算法达到了一个较好的跟踪效果，在一定程度上超越了现有算法，值得进一步改进与应用。

参考文献：

[1] Bolme D, Beveridge R J, Draper B A, *et al.* Visual object tracking using adaptive correlation filters [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2010: 2544-2550.

[2] Henriques. J F, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with Kernels [C]// Proc of European Conference on Computer Vision. 2012: 702-715.

[3] Henriques. J F, Caseiro R, Martins P, *et al.* High-speed tracking with kernelized correlation filters [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2015, 37 (3): 583-596.

[4] Danelljan M, Khan F S, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 1090-1097.

[5] 闫河，刘婕，杨德红，等．基于特征融合的粒子滤波目标跟踪新方法 [J]．光电子．激光，2014，25 (10)：1990-1999. (Yan He, Liu Jie, Yang Dehong, *et al.* A new particle filter object tracking algorithm based on feature fusion [J]. Journal of Optoelectronics. Laser, 2014, 25 (10): 1990-1999.)

[6] 张晓玄，高美凤．特征融合的核相关滤波目标跟踪 [J]．小型微型计算机系统, 2017, 38 (10) . (Zhang Xiaoxuan, Gao Meifeng. Kernel correlation filter target tracking based on feature fusion [J]. Journal of Chinese Computer Systems, 2017, 38 (10) .)

[7] 王春平，王暐，刘江义，等．基于色度饱和度-角度梯度直方图特征的尺度自适应核相关滤波跟踪 [J]．光学精密工程, 2016, 24 (9): 2293-2301. (Wang Chunping, Wang Wei, Liu Jiangyi, *et al.* Scale adaptive kernelized correlation filter tracking based on HHS-OG feature [J]. Optics and Precision Engineering, 2016, 24 (9): 2293-2301.)

[8] Danelljan M, Hager G, Khan F S, *et al.* Learning spatially regularized correlation filters for visual tracking [C]// Proc of IEEE International Conference on Computer Vision. 2016: 4310-4318.

[9] Danelljan M, Hager G, Khan F S, *et al.* Convolutional features for correlation filter based visual tracking [C]// Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2015: 621-629.

[10] 邢运龙，李艾华，崔智高，等．改进核相关滤波的运动目标跟踪算法 [J]．红外与激光工程, 2016, 45 (s1): 214-221. (Xing Yunlong, Li Aihua, Cui Zhigao, *et al.* Moving target tracking algorithm based on improved Kernelized correlation filter [J]. Infrared and Laser Engineering, 2016, 45 (s1): 214-221.)

[11] 赵翔东，石磊，刘飞．基于时空特征的实时相关滤波目标跟踪 [J]．传感器与微系统, 2016, 35 (8): 119-122. (Zhao Xiangdong, Shi Lei, Liu Fei. Real-time correlation filtering target tracking based on spatio-temporal feature [J]. Transducer and Microsystem Technologies, 2016, 35 (8): 119-122.)

[12] Ma Chao, Huang Jiabin, Yang Xiaokang, *et al.* Hierarchical convolutional features for visual tracking [C]// Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2015: 3074-3082.

[13] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4293-4302.

[14] Cui Zhen, Xiao Shengtao, Feng Jiashi, *et al.* Recurrently target-attending tracking [C]// Computer Vision and Pattern Recognition. 2016: 1449-1458.

[15] Danelljan M, Hager G, Khan F S, *et al.* Accurate scale estimation for robust visual tracking [C]// Proc of British Machine Vision Conference. 2014: 65. 1-65. 11.

[16] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: complementary learners for real-time tracking [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1401-1409.

[17] Danelljan M, Robinson A, Khan F S, *et al.* Beyond correlation filters: learning continuous convolution operators for visual tracking [C]// Proc of European Conference on Computer Vision. Springer, Cham, 2016: 472-488.

[18] 熊昌镇，赵璐璐，郭芬红．自适应特征融合的核相关滤波跟踪算法 [J]．计算机辅助设计与图形学学报, 2017, 29 (6): 1068-1074. (Xiong Changzhen, Zhao Lulu, Guo Fenhong. Kernelized correlation filters tracking based on adaptive feature fusion [J]. Journal of Computer-Aided Design and Computer Graphics, 2017, 29 (6): 1068-1074.)

[19] Wu Yi, Lim J W, Yang M H. Online object tracking: a benchmark [C]// Computer Vision and Pattern Recognition. 2013: 2411-2418.